

“Express Mail” Mailing No: EV327133247US
Date of Deposit: January 29, 2004

PATENT
Case No. 11336/677 (P01045US)

SYSTEM FOR SPEECH RECOGNITION

INVENTOR:

Ansgar Rinscheid

Attorney: Meredith Martin Addy
Registration No. 37,883
BRINKS HOFER GILSON & LIONE
P.O. Box 10395
Chicago, Illinois 60610
(312) 321-4200

SYSTEM FOR SPEECH RECOGNITION

INVENTOR:

5 **Ansgar Rinscheid**

PRIORITY CLAIM

[001] This application claims the benefit of European Application No. 03002067.1, filed January 29, 2003.

10 **BACKGROUND OF THE INVENTION**

1. Technical Field.

[002] This invention relates to a speech recognition system, and more particularly, to a speech recognition system that recognizes speech by comparing speech patterns to reference strings.

15 **2. Related Art.**

[003] Systems have been developed that recognize limited vocabularies spoken by specific individuals. When a statement is made, these systems compare spoken speech patterns to reference speech patterns. Using a template matching technique, probabilities are assigned to the speech patterns that most closely resemble the spoken patterns. Other speech 20 recognition systems apply complex algorithms to the spoken word. The application of complex algorithms may introduce delays that adversely affect a speech recognition system's performance.

[004] In a template matching speech system, speech recognition can be delayed by the large number of comparisons that occur when identifying a spoken word. The difficulty 25 of the search and the large number of evaluations may result in intolerable delays and complexity. Therefore, a need exists for a speech recognition system that understands the spoken word and minimizes objectionable delays.

SUMMARY

[005] This invention provides a speech recognition system that understands the spoken word for the purpose of receiving commands and data input from a speaker. The system may use a microphone to convert sound waves into electrical signals. The speech 5 recognition system includes a processor or controller that converts the microphone's output into data that represents a character string. A comparator compares the spoken character string to stored character strings. The stored character strings may be resident to a memory within or remote from the processor. The system recognizes the spoken word by first comparing a predetermined number of spoken characters to a predetermined number of 10 characters in the stored character strings. Based on the similarities and/or differences of the first comparison, all of the characters of selected character strings may be compared to all of the corresponding characters in the spoken character string. The selected character string that most closely matches the spoken character string may then be identified.

[006] An alternative speech recognition system compares a predetermined number 15 of spoken characters to a predetermined number of stored characters. Based on the similarities and/or differences, a number of character strings are selected. The number of characters in the selected strings to be compared to the spoken character string is then increased. Based on the similarities and/or differences in a second comparison, a second 20 number of character strings may be selected and the number of characters to be compared to the spoken character string through a third comparison is increased. The process continues until all of the spoken characters are compared to the corresponding stored characters of one or more selected character strings. Alternatively, the process may continue until a likelihood of a match exceeds a predetermined threshold.

[007] Other systems, methods, features and advantages of the invention will be, or 25 will become, apparent to one with skill in the art upon examination of the following figures and detailed description. It is intended that all such additional systems, methods, features and advantages be included within this description, be within the scope of the invention, and be protected by the following claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[008] The invention can be better understood with reference to the following drawings and description. The components in the figures are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention. Moreover, in the figures, like referenced numerals designate corresponding parts throughout the different views.

- [009] Figure 1 is a block diagram of a speech recognition system.
- [010] Figure 2 is a flow diagram of a process of speech recognition.
- [011] Figure 3 is a flow diagram of an alternative process of speech recognition.
- [012] Figure 4 is a block diagram of a memory.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[013] The speech recognition system may recognize a spoken word to receive commands and obtain data from a system. The system may be used to transmit, store, manipulate or even create speech. The speech recognition system minimizes delays while maintaining speech recognition quality.

[014] Figure 1 is a block diagram of a speech recognition system. The system may include a microphone 101, a digital signal processor 102, a comparator 103, a control unit 104 and a memory 105. The microphone 101 receives sound and converts sound waves into analog or digital signals which are directed to the processor 102. The processor 102 digitizes the audio signal into spoken character strings that may have variable length. The character strings may contain any sequence of elements from a given character set, such as letters, and are transmitted to the comparator 103 which is in communication with the controller or a computer unit 104 and the memory 105. The comparator 103 compares these character strings to corresponding character strings from memory 105. The control unit 104 takes the result of the comparison and calculates probabilities for a match in an attempt to identify the character string closest to that representing the sound. Ultimately, the character string in memory 105 having the highest probability for a match may be recognized as correlating to the speech signal and that patterned, character string will be processed as the one representing the sound.

[015] The character string in memory having the highest probability of matching the speech is determined by the flow diagram in Figure 2. Once the speech signal has been processed by the digital signal processor 102 and the comparator 103 receives the character string, just a small group of characters, defined by a predetermined number N of characters in the spoken character string 201, will be compared against small groups of characters having the same number N of characters from each set of corresponding character strings 202 in memory 105. As each comparison 203 is made in the comparator 103, the control unit 104 calculates 204 the probability of a match for a possible identification between each of the corresponding character strings in memory 105 and the spoken character string. The control unit 104 has a predetermined minimum value for the probability of a match. For the group of stored character strings that exceed the minimum value 205, those strings will be selected 206 to advance to the next step of the process.

[016] In the next step, the remaining characters from the group of stored character strings will be added 207 to the first N characters used in the first comparison. Also, the remaining characters for the spoken character string will be added 208 to those N characters that were part of the first comparison step 203. At this point, another comparison 209 between all of the characters of the pattern character strings or the selected group of stored characters and all of the characters in the spoken character string may be made in the comparator 103. Using the comparison results, the probabilities for a match are once again calculated 210 by the control unit 104 for each of the selected character strings. The selected character string with the highest probability for a match is determined 211 and that selected character string is the one that is recognized as identifying the sound that was received by the speech recognition system.

[017] Alternatively, rather than adding the remaining characters all at once to the first group of characters in the applicable character strings, the remaining characters may be added to the character strings in incremental steps. In this process, as shown in Figure 3, one character, or a few characters, may be added to the selected character strings and the comparison 309, calculation 304 and selection steps 305 and 306 may be repeated in a step-by-step fashion. Again, one character, or a few characters, may be added to the character strings 307 and 308 in subsequent cycles.

[018] Just as it was performed previously, after each comparison 304, the control unit 104 determines 305 which of the newly selected group of stored character strings have

the highest probability for identifying the speech. Those abbreviated, selected character strings, that exceed the predetermined value of probability for that cycle, may be selected 306 once again, and one character, or a few more characters, may be added 307 to these newly selected character strings for further comparison. This process is repeated until all of the 5 characters have been added 310 to the character string or one from the group of stored character strings from memory 105 has been recognized 311 as attaining the highest probability for identifying the sound. Thus, the sound will be recognized 312 and represented in any further processing by that selected character string.

[019] The speech recognition system can best be understood by means of an 10 illustrated example. Figure 4 shows a layout of the memory 405 with a predetermined reference of M pattern character strings for recognizing speech. The memory 405 is shown as a matrix memory with 15 lines and 11 columns. Each line 1-15 represents a pattern character string. M in this case equals 15.

[020] As an example, the speech or sound is received by the microphone 101 and 15 formed into the character string ABCXOPQRSTX by the digital signal processor 102. This character string represents the speech or sound that needs to be identified. In the first step of this method, it is assumed for purposes of illustration, that N will equal 4 so that the first 4 characters of the stored character strings 403 and the first 4 characters of the spoken or sound character string 401 to be identified are taken into consideration. Thus, the first four 20 characters, ABCX, of the spoken or sound character string 401 are compared to the first four characters of the pattern character string 403 in the fifteen lines of memory 205. In the pattern character strings 1, 6, 10, 11 and 12, hits exist. In the pattern character strings 1, 10, 11 and 12, three of four characters correspond to the character string to be identified.

[021] Simplistically, the probabilities are calculated as a quotient from the number 25 of hits in each comparison to the total number of characters in a stored character string that were compared. For this example, the minimum value for a probability of a match, indicating an identification might be made, is assumed to be 75 %. In the corresponding string of characters in line 6 and in line 15 only one character corresponds to the speech character string. The probability for a match in these lines is therefore $\frac{1}{4}$ or 25 %. The corresponding 30 strings of characters of lines 1, 6, 10, 11, 12 and 15 had hits, but only the pattern strings of characters of lines 1, 10, 11 and 12 will be selected for the last process step because they meet the minimum probability for a match of 75%. In the remaining character strings, the

comparison shows that no characters correspond to the spoken character string. Therefore, these lines would not be selected for further processing.

[022] All of the remaining characters for the selected character strings 1, 10, 11 and 12 are added to the original four characters that were analyzed and are now taken into consideration in the following comparison step. Likewise, the remaining characters are added to the original N (four) characters of the speech character string and it is compared to the pattern character strings 1, 10, 11 and 12. For the character string in line 10, six more hits are obtained, whereas in the character string in line 12 only one more hit occurred. In the selected character strings of lines 1 and 11, further hits did not occur. Referring to the pattern character string of line 10, the probability for a match is 9/11, or 81.81%. The probability for identity is 4/11 or 36.36% for the pattern character string of line 12. Thus, the pattern character string in line 10 is recognized as the character string that identifies the speech.

[023] In the alternative method, not all of characters of the pattern character string and the speech character string are taken into consideration all at once. Rather, the number of characters to be taken into consideration in the follow-on steps is increased step-by-step. For this example, one additional character is added step-by-step. After each comparison, calculation and selection, those pattern character strings that are selected will have one character added to the character string until all of the characters in columns 5 to 11 for the selected character strings are considered. It is clear to see that pattern character string 10 would clearly be the only one remaining after the first step because it is the only one increasing in probability for each of the first six steps.

[024] Although the number of comparison cycles increase, the number of pattern character strings analyzed will be reduced as illustrated in the above example, thus reducing the time for the comparison and calculation steps. In the end, the selected character string having the highest probability is the character string that identifies the speech. In the above examples, the same corresponding character string was recognized using both methods, just as it should have been.

[025] Further, the number of characters added to the character strings could have been more than just one at a time. Two or more characters may have been added to the strings in each step, thus possibly increasing the speed at which the identification would occur.

[026] In the system described above, a few characters in all of the stored character strings may be compared against a few characters in the spoken character string in the initial step. The time for locating and identifying the proper character string may be significantly reduced. In this way, the system may target and focus toward the most probable location for the stored, patterned character string. The time saved, when compared to a complete comparison of all the characters in each pattern character string, may be determined by the selection of that first grouping of characters in the character strings. The smaller the number of characters, the less time it takes to identify the speech.

[027] While various embodiments of the invention have been described, it will be apparent to those of ordinary skill in the art that many more embodiments and implementations are possible within the scope of the invention. Accordingly, the invention is not to be restricted except in light of the attached claims and their equivalents.